PYTHON_WORKSHEET_1

Answer (1) – (A) True

Answer (2) – (A) Central limit theorem

Answer (3) – (B) Modeling bounded count data

Answer (4) – (d) All of the mentioned

Answer (5) – (C) Poisson

Answer (6) – (B) false

Answer (7) – (B) Hypothesis

Answer (8) – (A) 0

Answer (9) – (C) Outliers cannot conform to the regression relationship

Subjectives

1. **What do you understand by the term Normal Distribution?**

- A normal distribution or Gaussian distribution refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency. Thus, a bell-shaped curve is formed.

- A normal distribution is a statistical phenomenon representing a symmetric bell-shaped curve. Most values are located near the mean; also, only a few appear at the left and right tails.
- It follows the empirical rule or the 68-95-99.7 rule.
- The mean, median, and mode are equal and the mean and standard deviation of the function are 0 and 1, respectively.
- This mathematical function has two key parameters:
  The mean (μ) and the standard deviation (σ).

2. **How do you handle missing data? What imputation techniques do you recommend?**

Missing data can be dealt with in a variety of ways.:-

- The most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you.
- Your application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea.
- Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

❖ I am recommending three imputation techniques: -

1) regression imputation
2) predictive mean matching
3) hot deck imputation

3. **What is A/B testing?**

- A/B testing in its simplest sense is an experiment on two variants to see which performs better based on a given metric. Typically, two consumer groups are exposed to two different versions of the same thing to see if there is a significant difference in metrics like sessions, click-through rate, and/or conversions.

- Using the visual above as an example, we could randomly split our customer base into two groups, a control group and a variant group. Then, we can expose our variant group with a red website banner and see if we get a significant increase in conversions. It's important to note that all other variables need to be held constant when performing an A/B test.

4. **Is mean imputation of missing data acceptable practice?**

- No, mean imputation is typically considered terrible practice since it ignores feature correlation

5. **What is linear regression in statistics?**

- Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

6. **What are the various branches of statistics?**

There are three branches of statistics:

- Data collection
- Descriptive statistics
- Inferential statistics